DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

National Institutes of Health
National Human Genome
Research Institute
31 Center Drive MSC 2152
Building 31, Room 4B09
Bethesda, MD 20892-2152
(301) 496-0844
FAX (301) 402-0837

COPY

February 28, 2000

CONFIDENTIAL

J. Craig Venter, Ph.D.
Mr. Tony White
Arnold Levine, Ph.D.
Mr. Paul Gilman
Celera Genomics Corporation
45 West Gude Drive
Rockville, Maryland 20850

Dear Dr. Venter, Mr. White, Dr. Levine and Mr. Gilman:

We appreciated the chance to meet with you on December 29, 1999, to discuss possible grounds for a collaboration between the public Human Genome Project (HGP) and Celera Genomics, on deriving the sequence of the human genome. Prior to the December 29 meeting, there had been numerous discussions about a possible model for collaboration, most notably those among Eric Lander, Craig Venter, Harold Varmus, and Arnie Levine. The five largest genome centers (Baylor, the JGI, Sanger, Washington University, and the Whitehead Institute, referred to as the G5) had been fully briefed on those discussions, and had asked the four of us to represent the public HGP in any negotiations. Other members of the international sequencing consortium were also briefed in general on these developments. We had prepared a proposed statement of "Shared Principles" (enclosed) which we thought accurately reflected the conclusions of the earlier conversations with Celera. From the previous contacts, we did not expect these to be particularly controversial, and were thus disappointed to learn the day before the meeting that some of those principles might not be shared by Celera. Nonetheless, we attempted to negotiate on December 29 in good faith.

A number of important points of agreement between ourselves and Celera were reached on December 29. These included the conclusion that an active collaboration to merge the two data sets could produce a substantially better product, since it would allow bilateral sharing of electrophoretic traces and a joint effort to resolve discrepancies. At a technical level, this kind of data sharing did not seem to present a substantial hurdle in the implementation of a joint effort.

However, several fundamental differences emerged. These are outlined below:

1) Since its inception, the public HGP has been committed to producing human genomic sequence for all to use without any barriers or restrictions, in the expectation that this policy would maximize the benefit to humankind. That policy was explicitly codified in 1996. If a collaboration were initiated, however, Celera indicated they would expect exclusive commercial rights of distribution of the merged product, most likely even beyond the current projected date of completion of the

human genome sequence by the public consortium in 2003 (2005 was mentioned). While we understand Celera's business imperatives, the extent of this claim seemed inappropriate to us. The product of a sequence data merge between the HGP and Celera in the summer of 2000 would reflect the collection of shotgun data at roughly 10x redundancy, and would not be of equivalent quality to the finished product expected from the HGP in 2003. Many sequence gaps and regions of ambiguity would be present in the proposed Celera/HGP merged product. By current plans, if no collaboration were initiated, the public HGP expects to have completed shotgun coverage on available genomic clones by no later than December 2000. This would argue that the period of exclusive rights to commercial distribution of the merged data set by Celera should be limited to no more than 6 - 12 months after the initiation of the collaboration.

2) In view of the public HGP's commitment to finishing the human sequence, it would be expected that ongoing efforts would be needed after the merge to close gaps and clear up ambiguities. The design of such finishing strategies would optimally make full use of the merged data set. On December 29, however, Celera argued that the output of any further improvements in the sequence that resulted from the use of the merged data set would also fall under their exclusive rights for commercial distribution. Such a restriction would put the public HGP in an ongoing position of producing data that was less than fully accessible, which would be inconsistent with the internationally agreed principle of open access. To avoid this, the public HGP might actually be forced to continue to collect additional independent and redundant shotgun data in parallel with the collaborative effort, although this position was apparently also unacceptable to Celera. Given Celera's stated plans to make their own version of the data available to the academic community, even in the absence of a joint effort, it is not clear why or how Celera would expect to prevent the public HGP from utilizing the complete merged data set in directing finishing efforts.

3) The public HGP was also concerned by Celera's stated desire to have their proposed commercial rights over the joint product extend beyond databases to experimental applications, such as the construction of genome chips, large primer sets, or applications to proteomics and analysis of regulatory sequences. We were not previously aware of the intention for those commercial rights to extend beyond databases. While establishing a monopoly on commercial uses of the human genome sequence may be in Celera's business interests, it is not in the best interests of science or the general public.

4) While both parties agreed that wide distribution via both a DVD and the internet was desirable to facilitate access to the sequence, there appeared to be substantial differences in how this might be implemented. The public HGP feels that the sequence, along with jointly derived annotation, must be available from one or more noncommercial sites, whereas Celera wishes any internet access to be exclusively through a Celera portal.

5) We were concerned about Celera's possible intention to publish the results of their merge of the public data and their own data, in the event that no agreement is reached. Although the public effort has explicitly made preliminary sequence available for all to use for both academic and commercial research purposes, the public consortium reasonably reserves the right to be the first to publish its own data in a peer-reviewed journal. Publication of other groups' primary data without consent is considered to be a breach of scientific ethics. In addition, it is counter to accepted scientific practice for authors to present results without having examined the primary data.

These are substantive issues. We left the December 29 meeting grateful that the exchange had been so open and wide-ranging, but uncertain whether Celera's stated positions were hardened or still open to negotiation. At the conclusion of the meeting, all parties agreed that a follow-up meeting would probably be needed.

One of us (FSC) attempted to reach Dr. Venter to set up a follow-up meeting, but was directed instead to Mr. White. In a phone call on January 22, Mr. White reiterated the points outlined above, and indicated that a follow-up meeting would only make sense if the public consortium agreed to grant Celera commercial protection for the sequence in their database well beyond the time it would have taken the public HGP to generate an equivalent product. Mr. White stated that he expected commercial protection for at least 3 to 4 years, and preferably longer. He also held fast to the demand that finishing efforts by the public consortium, carried out after the merge, would need to be under the same restrictions. When it was pointed out that this was particularly puzzling, since if there were no collaboration the HGP could presumably use Celera's proposed DVD to aid finishing and then place those additional new reads in public databases, Mr. White indicated that they were rethinking their DVD plans, and that the sequence data might only be available on their web site, perhaps with some restriction on this sort of finishing activity by others.

When asked about the extent to which Celera would see their restrictions applying to uses of the genome sequence other than commercial databases, Mr. White reiterated that Celera would expect chip companies to pay a license for this use. As to other genome-wide uses of the human sequence, he said that Celera has not yet defined which types of experiments would require a license, but left the door open to that being broader than just chips.

When asked whether there was some flexibility in the positions outlined by Celera on Dec. 29, Mr. White did not encourage that view.

Despite this discouraging interaction, one of us (FSC) has been attempting for more than two weeks to reach Dr. Venter to see whether there are any remaining grounds for pursuing the collaborative model. After several unanswered phone calls and e-mails outlining the reason for the attempted contact, Dr. Venter's assistant indicated that he is too busy with other matters right now to discuss this until some time after March 6.

A clear conclusion seems to arise from Celera's actions on and subsequent to December 29: that there is no real interest on the part of Celera in continuing to pursue this particular collaborative model. This is disappointing, since much work has gone into exploring such an option, and we had been hopeful in light of the success of the collaborative effort on the Drosophila sequence that something similar could be worked out for the human project.

Continued ambiguity is not conducive to early completion of the human genome sequence. Therefore, unless Celera indicates in the next week (by March 6, 2000) that the conclusions of this letter are substantively incorrect, we will conclude that the initial proposal whereby the data from the public HGP and Celera are collaboratively merged is no longer workable.

We stand ready and willing to engage in further conversations at any time.

With best regards,

Sincerely yours,

Francis S. Collins, M.D., Ph.D.

Harold Varmus, M.D.

Martin Bobrow, CBE DSC FRCP

Robert H. Waterston, M.D., Ph.D.

CONFIDENTIAL

December 28, 1999

SHARED PRINCIPLES

There seems to be general agreement that:

o The public HGP and Celera Genomics each have the capacity to generate substantial coverage of the sequence of the human genome.

o Humankind will be better served if we can find a viable way to join forces to produce a better product in a more timely fashion.

o The methods being used (clone-based and whole-genome shotgun) are, in fact, complementary. They provide much opportunity for cross-checking.

o A collaboration would offer the opportunity for joint optimization of experimental strategy and analytical methods.

o The current antagonism and excessive competition should be replaced with a more collaborative spirit.

o The public HGP is committed to the complete sequence of the human genome being freely available in the public domain.

o The public HGP understands that Celera Genomics will be making its consensus sequence of the human genome broadly available.

o Celera Genomics is able to see assembled data generated by the public labs and is free to use it for its proprietary databases, but a scientific publication that combines substantial data from both sources should, according to accepted scientific practice, be a joint publication involving authors from both groups.

The outline of a collaborative effort might be as follows:

1. Joint Publication.

The two sides would agree to joint publication of a collaborative paper reporting joint analysis of both the public and Celera data, co-authored by appropriate scientists from both sides. An additional paper reporting analysis of the public data from the public HGP labs might also be simultaneously submitted.

The collaborative paper would necessarily involve that co-authors from the public side have sufficient access to the Celera data and any joint analysis during the preparation of a

paper to be able to comfortably sign their names to a paper.

There is some concern about the appearance of public HGP scientists having access to Celera's data before it is made public. The best solution might be to limit the period of access, perhaps to 90 days before submission of a paper.

The most realistic timeframe would be to aim for the paper(s) to appear near the end of 2000 (perhaps in Science), which would require submission in the late summer or fall and would involve collaboration beginning in the late spring or summer.

It is important that the public consortium's participation in this collaboration does not contribute directly to development of intellectual property for Celera, whether in terms of sequence or SNPs.

2. Data Release.

The public HGP will continue to release their data immediately.

Celera and other companies would continue to be free to use these data to create and sell databases.

Upon publication of the papers, the human sequence produced from the analysis of the public and Celera data will also be released in accessible "media". The ideal medium (because of its universal accessibility) would be an internet-accessible database. Access solely through Celera's database has substantial disadvantages from the public consortium's perspective.

3. Public Relations

Consideration will be given to preparing a brief statement indicating that the possibility of collaboration is being explored. During the negotiations no information will be provided to the press by either party.

If agreement is reached, the two sides will jointly prepare a document and press releases describing the rationale for and nature of the collaboration. Both sides will make all efforts to support this agreement. This includes honest and straightforward responses to scientific inquiries, but avoiding disparaging the other party, sowing discord or undermining the collaborative spirit. It will require a generosity of spirit in acknowledging the legitimate and important role of both parties.